Adverse Event Prediction Using Graph-Augmented Temporal Analysis

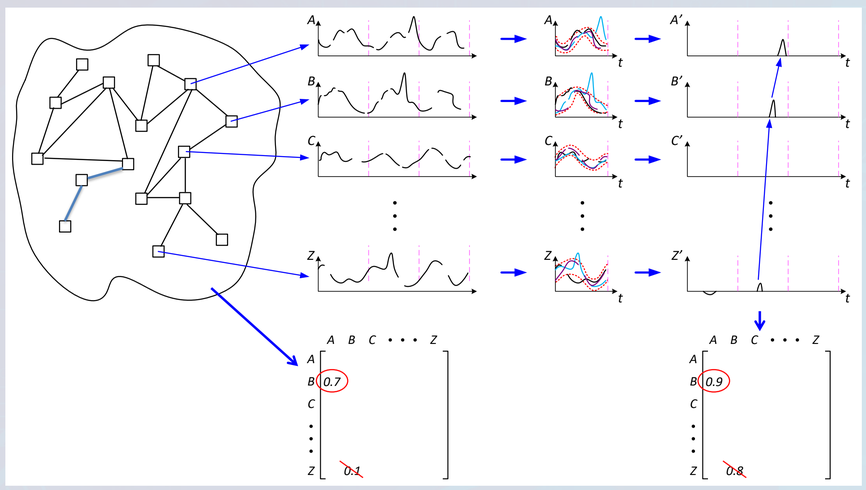
Sandia National Laboratories
Randy Brost, Vitus Leung, Hamilton Link, Cindy Phillips, Andrea Staid
Sandia National Laboratories, Albuquerque, NM 87185

Problem

- Data streams are voluminous and complex.
- Hard to understand event chains leading up to events.
- Prior work: Spatiotemporal search for specified patterns.
- This project seeks to find unknown temporal relationships.

Approach

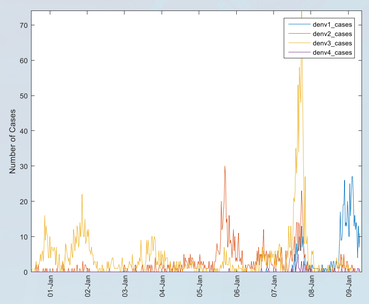
Find precursors via temporal and graph analysis.



A domain has multiple data streams logged at various points. Temporal analysis identifies potential correlations, which are either reinforced or deprecated by graph information.

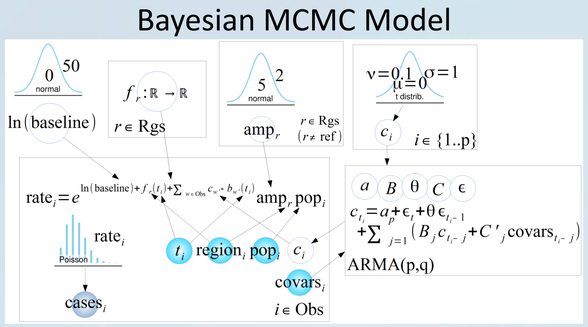
Example Problems:

1. Dengue Fever Outbreaks:



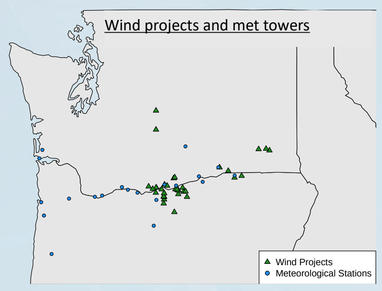
Data: Dengue cases, weather…
Goal: What precedes an outbreak?

Bayesian MCMC Model



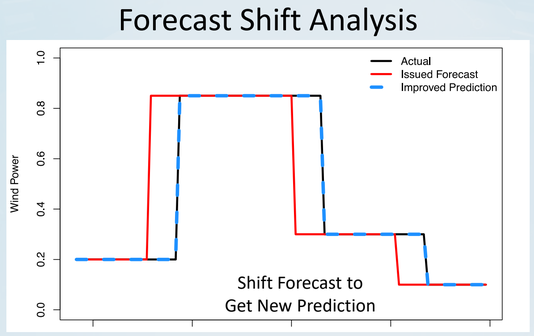
Method: Construct Bayesian model relating context, weather, and Dengue cases. Characterize priors via Monte Carlo analysis. Find parameters connected to outbreaks.

2. Wind Power Analysis:



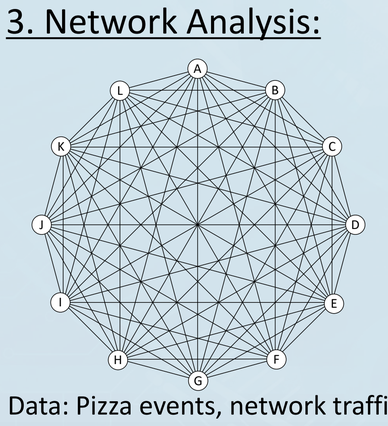
Data: Wind power forecast, actual…
Goal: Can we predict forecast errors?

Forecast Shift Analysis



Method: Identify adjustments to forecast based on recent, local actual performance.

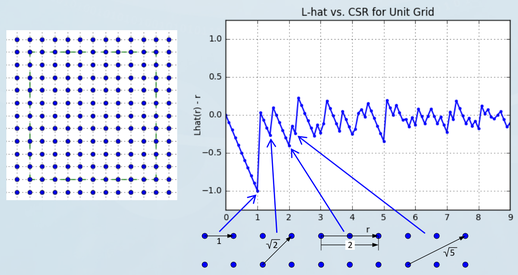
3. Network Analysis:



Data: Pizza events, network traffic.
Goal: Who orders the pizza?

Ripley's K Function

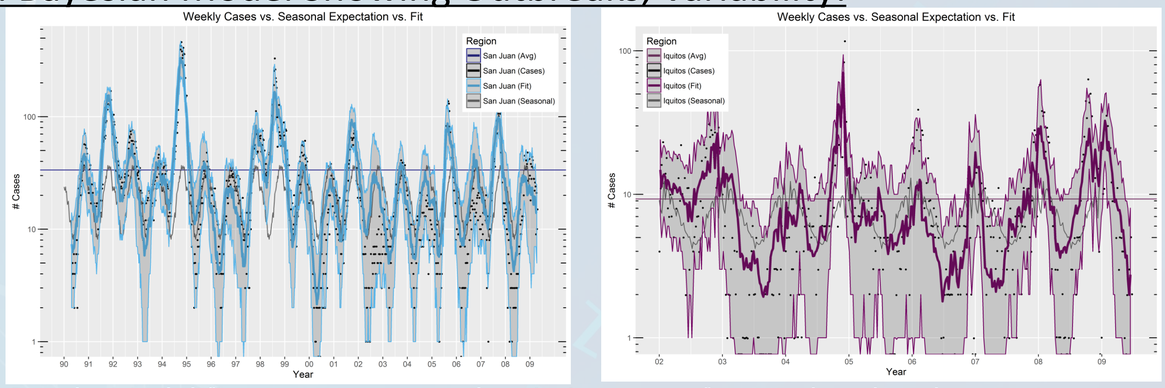
$$\hat{K}(r) = \hat{\lambda}^{-1} \sum_{i}^{N} \sum_{j \neq i}^{N} \frac{I[d_{ij} < r]}{N} \qquad \hat{L}(r) = \sqrt{\frac{\hat{K}(r)}{\pi}}$$



Method: Identify structures in spatial and temporal data that deviate from random.

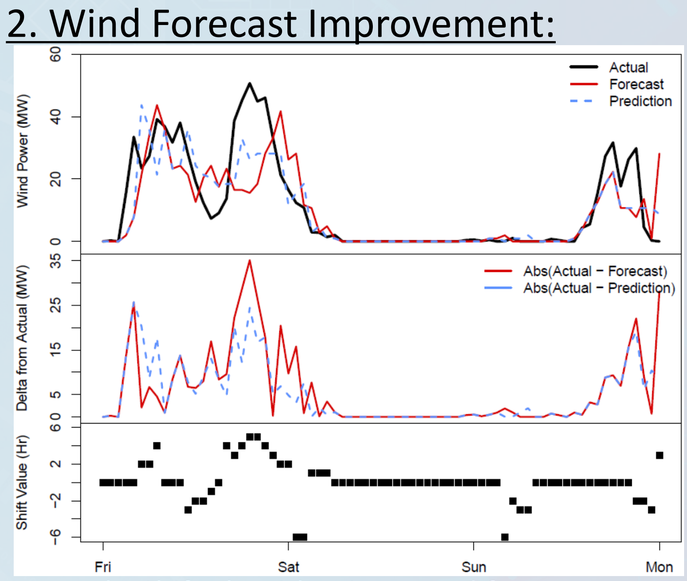
Results

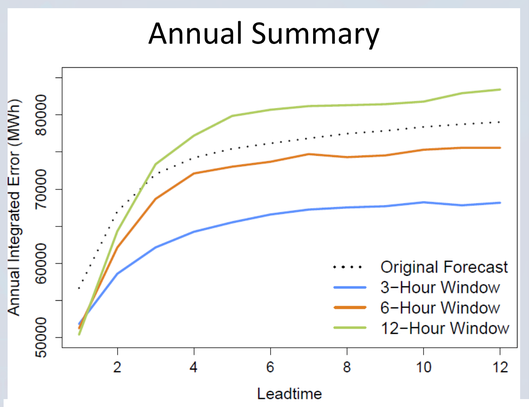
1. Bayesian Model Showing Outbreaks, Variability:



Result: Model fit captures seasonal variation, out-of-normal outbreaks.
Next: Assess environmental influence, real-time prediction.

2. Wind Forecast Improvement:

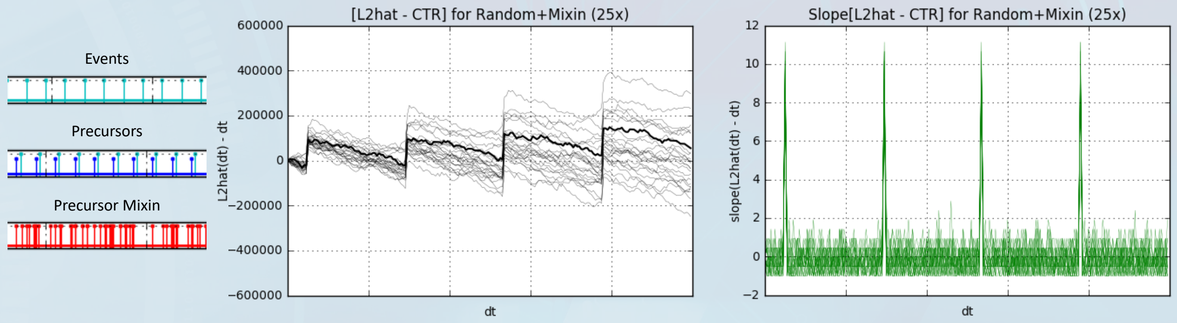


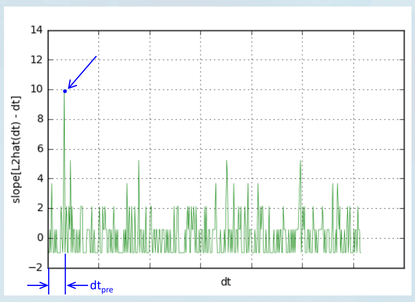


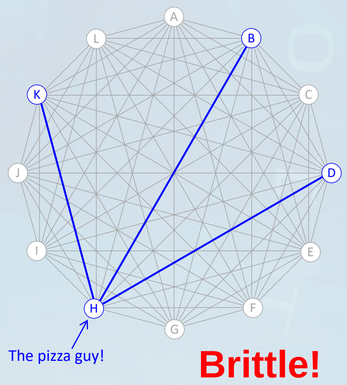
Data courtesy Bonneville Power Administration.

Result: Shift algorithm improved forecast accuracy by 9,873 MWh/year (5.1%).
Next: Improved algorithms, geographic coupling.

3. Network Precursor Analysis (synthetic data):

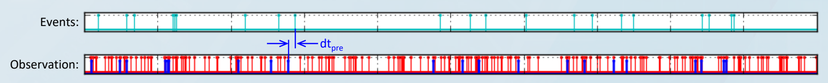






Brittle!

Identify precursors in observation:



Result: Modified K-function analysis found precursor network traffic within pure synthetic noise, given prefect precursor lead time.
Next: Add variable lead time, noise, real data. Seek robust method.

Significance

Potential impact:

- New algorithms for solving spatiotemporal analysis problems.
- Supports forensics, warning, relationship identification.
- Multiple potential national security applications.

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND 2017-8559 D

Sandia National Laboratories